

10-02-00

A

UTILITY PATENT APPLICATION TRANSMITTAL

Attorney Docket No.

50325-0106

(New Nonprovisional Applications Under 37 CFR § 1.53(b))

TO THE COMMISSIONER FOR PATENTS:

Transmitted herewith is the patent application of () application identifier or (X) first named inventor, Arthur Zavalovsky, et al., entitled MARKING NETWORK DATA PACKETS WITH DIFFERENTIATED SERVICES CODEPOINTS BASED ON NETWORK LOAD, for a(n):

(X) Original Patent Application.

() Continuing Application (prior application not abandoned):

() Continuation () Divisional () Continuation-in-part (CIP)
of prior application No: _____ Filed on: _____

() A statement claiming priority under 35 USC § 120 has been added to the specification.

Enclosed are:

(X) Specification 30 Total Pages; (X) Drawing(s); 5 Total Sheets; (X) Cover Sheet 1 Page() Oath or Declaration: 3 Pages

() A Newly Executed Combined Declaration and Power of Attorney:

() Signed.

(X) **Unsigned.**

() Partially Signed.

() A Copy from a Prior Application for Continuation/Divisional (37 CFR § 1.63(d)).

() Incorporation by Reference. The entire disclosure of the prior application, from which a copy of the oath or declaration is supplied, is considered as being part of the disclosure of the accompanying application and is hereby incorporated herein by reference.

() Signed Statement Deleting Inventor(s) Named in the Prior Application. (37 CFR § 163(d)(2)).

() Power of Attorney.

(X) Return Receipt Postcard.

() Associate Power of Attorney.

() A Check in the amount of \$ _____ for the Filing Fee.

() Preliminary Amendment.

() Information Disclosure Statement and Form PTO-1449.

() A Duplicate Copy of this Form for Processing Fee Against Deposit Account.

() A Certified Copy of Priority Documents (if foreign priority is claimed).

() Statement(s) of Status as a Small Entity.

() Statement(s) of Status as a Small Entity Filed in Prior Application, Status Still Proper and Desired.

() Other: _____

CLAIMS AS FILED

FOR	NO. FILED	NO. EXTRA	RATE	FEE
Total Claims	14	0	\$18.00	\$ 0.00
Independent Claims	5	2	\$78.00	\$ 156.00
Multiple Dependent Claims (if applicable)				\$0.00
Assignment Recording Fee				\$0.00
Basic Filing Fee				\$690.00
Total Filing Fee				\$ 846.00

Charge \$ _____ to Deposit Account _____ pursuant to 37 CFR § 1.25. At any time during the pendency of this application, please charge any fees required or credit any overpayment to this Deposit Account.

Respectfully submitted,

By: _____

Carl L. Brandt, Reg. No. 44,555

Date: September 29, 2000

Correspondence Address:

Hickman Palermo Truong & Becker, LLP
1600 Willow Street
San Jose, California 95125-5106
Telephone: (408) 414-1080
Facsimile: (408) 414-1076

I hereby certify that this is being deposited with the U.S. Postal Service "Express Mail Post Office to Addressee" service under 37 CFR § 1.10 on the date indicated below and is addressed to:

Commissioner for Patents
Box Patent Application
Washington, D.C. 20231

By: _____

Typed Name: Casey Moore

Express Mail Label No.: EL624353785US

Date of Deposit: September 29, 2000

UNITED STATES PATENT APPLICATION

FOR

MARKING NETWORK DATA PACKETS WITH DIFFERENTIATED SERVICES
CODEPOINTS BASED ON NETWORK LOAD

INVENTORS:

ARTHUR ZAVALKOVSKY
ITZHAK PARNAFES
SHAI MOHABAN

PREPARED BY:

HICKMAN PALERMO TRUONG & BECKER, LLP
1600 Willow Street
San Jose, CA 95125
(408) 414-1080

EXPRESS MAIL CERTIFICATE OF MAILING


"Express Mail" mailing label number EL624353785US

Date of Deposit September 29, 2000

I hereby certify that this paper or fee is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 CFR 1.10 on the date indicated above and is addressed to the Assistant Commissioner for Patents, Washington, D.C. 20231.

Casey Moore

(Typed or printed name of person mailing paper or fee)



(Signature of person mailing paper or fee)

0967580 096000

MARKING NETWORK DATA PACKETS WITH DIFFERENTIATED SERVICES
CODEPOINTS BASED ON NETWORK LOAD

FIELD OF THE INVENTION

5 The present invention relates generally to computer networks, and more specifically, to a method and apparatus for marking data packets with differentiated services codepoints based on network traffic load.

BACKGROUND OF THE INVENTION

10 A computer network typically comprises a plurality of interconnected entities that transmit ("source") or receive ("sink") data frames. A common type of computer network is a local area network ("LAN") that generally comprises a privately owned network within a single building or campus. LANs employ a data communication protocol (LAN standard) such as Ethernet, FDDI, or Token Ring, that defines the functions performed by
15 the data link and physical layers of a communications architecture (i.e., a protocol stack), such as the Open Systems Interconnection (OSI) Reference Model. In many instances, multiple LANs may be interconnected by point-to-point links, microwave transceivers, satellite hookups, etc., to form a wide area network ("WAN"), metropolitan area network ("MAN") or Intranet. These internetworks may be coupled through one or more gateways
20 to the global, packet-switched internetwork generally known as the Internet or World Wide Web (WWW).

Each network entity preferably includes network communication software, which may operate in accordance with Transport Control Protocol/Internet Protocol (TCP/IP). TCP/IP generally consists of a set of rules defining how entities interact with each other.
25 In particular, TCP/IP defines a series of communication layers, including a transport layer and a network layer. At the transport layer, TCP/IP includes both the User Data Protocol (UDP), which is a connectionless transport protocol, and TCP, which is a reliable,

connection-oriented transport protocol. When a process at one network entity wishes to communicate with another entity, it formulates one or more messages and passes them to the upper layer of the TCP/IP communication stack. These messages are passed down through each layer of the stack where they are encapsulated into packets and frames. Each layer also adds information in the form of a header to the messages. The frames are then transmitted over the network links as bits. At the destination entity, the bits are re-assembled and passed up the layers of the destination entity's communication stack. At each layer, the corresponding message headers are stripped off, thereby recovering the original message that is handed to the receiving process.

One or more intermediate network devices are often used to couple LANs together and allow the corresponding entities to exchange information. For example, a bridge may be used to provide a "bridging" function between two or more LANs. Alternatively, a switch may be utilized to provide a "switching" function for transferring information, such as data frames or packets, among entities of a computer network. Typically, the switch is a computer having a plurality of ports that couple the switch to several LANs and to other switches. The switching function includes receiving data frames at a source port and transferring them to at least one destination port for receipt by another entity. Switches may operate at various levels of the communication stack. For example, a switch may operate at Layer 2, which in the OSI Reference Model, is called the data link layer, and includes the Logical Link Control (LLC) and Media Access Control (MAC) sub-layers.

Other intermediate devices, commonly known as routers, may operate at higher communication layers, such as Layer 3, which in TCP/IP networks corresponds to the Internet Protocol (IP) layer. Conventionally, IP data packets include a corresponding header that contains an IP source address and an IP destination address. Routers or Layer 3 switches may re-assemble or convert received data frames from one LAN standard (e.g., Ethernet) to another (e.g., Token Ring). Thus, Layer 3 devices are often used to

interconnect dissimilar subnetworks. Some Layer 3 intermediate network devices may also examine the transport layer headers of received messages to identify the corresponding TCP or UDP port numbers being utilized by the corresponding network entities. Many applications are assigned specific, fixed TCP and/or UDP port numbers in accordance with Request For Comments (RFC) 1700. For example, TCP/UDP port number 80 corresponds to the Hypertext Transport Protocol (HTTP), while port number 21 corresponds to File Transfer Protocol (FTP) service.

ALLOCATION OF NETWORK RESOURCES

A process executing at a network entity may generate hundreds or thousands of traffic flows that are transmitted across a network. Generally, a traffic flow is a set of messages (frames and/or packets) that typically correspond to a particular task, transaction or operation (e.g., a print transaction) and may be identified by various network and transport parameters, such as source and destination IP addresses, source and destination TCP/UDP port numbers, and transport protocol.

The treatment that is applied to different traffic flows may vary depending on the particular traffic flow at issue. For example, an online trading application may generate stock quote messages, stock transaction messages, transaction status messages, corporate financial information messages, print messages, data backup messages, etc. A network administrator may wish to apply a different policy or service treatment ("quality of service" or "QoS") to each traffic flow. In particular, the network administrator may want a stock quote message to be given higher priority than a print transaction. Similarly, a \$1 million stock transaction message for a premium client should be assigned higher priority than a \$100 stock transaction message for a standard customer.

Computer networks include numerous services and resources for use in moving traffic throughout the network. For example, different network links, such as Fast Ethernet, Asynchronous Transfer Mode (ATM) channels, network tunnels, satellite links, etc., offer unique speed and bandwidth capabilities. Additionally, the intermediate

devices also include specific resources or services, such as number of priority queues, filter settings, availability of different queue selection strategies, congestion control algorithms, etc.

Individual frames or packets can be marked so that intermediate devices may treat them in a predetermined manner. For example, the Institute of Electrical and Electronics Engineers (IEEE) describes additional information for the MAC header of Data Link Layer frames in Appendix 802.1p to the 802.1D bridge standard.

FIG. 1A is a partial block diagram of a Data Link frame 100 that includes a MAC destination address (DA) field 102, a MAC source address (SA) field 104 and a data field 106. According to the 802.1Q standard, a user_priority field 108, among others, is inserted after the MAC SA field 104. The user_priority field 108 may be loaded with a predetermined value (e.g., 0-7) that is associated with a particular treatment, such as background, best effort, excellent effort, etc. Network devices, upon examining the user_priority field 108 of received Data Link frames 100, apply the corresponding treatment to the frames. For example, an intermediate device may have a plurality of transmission priority queues per port, and may assign frames to different queues of a destination port on the basis of the frame's user priority value.

FIG. 1B is a partial block diagram of a Network Layer packet 120 corresponding to the Internet Protocol. Packet 120 includes a type_of_service (ToS) field 122, a protocol field 124, an IP source address (SA) field 126, an IP destination address (DA) field 128 and a data field 130. The ToS field 122 is used to specify a particular service to be applied to the packet 120, such as high reliability, fast delivery, accurate delivery, etc., and comprises a number of sub-fields. The sub-fields may include a 3-bit IP precedence (IPP) field and three one-bit flags that signify Delay, Throughput, and Reliability. By setting the flags, a device may indicate whether delay, throughput, or reliability is most important for the traffic associated with the packet.

FIG. 1C is a partial block diagram of a Transport Layer packet 150 that preferably includes a source port field 152, a destination port field 154, and a data field 156, among others. Fields 152, 154 preferably are loaded with the TCP or UDP port numbers that are utilized by corresponding network entities.

5 DIFFERENTIATED SERVICES

Currently, a Differentiated Services (DS) model is under development by the Internet Differentiated Services Working Group of the Internet Engineering Task Force (IETF). The main idea behind DS is the classification and possibly conditioning of traffic at network boundaries. The classification operation entails the assignment of network
10 traffic to behavioral aggregates. The behavioral aggregates define a collection of packets with common characteristics that determine how they are identified and treated by the network.

To achieve the classification, the Internet Differentiated Services Working Group has proposed replacing the ToS field 122 of Network Layer packets 120 with a one-octet
15 differentiated services (DS) field 132, which is assigned a differentiated services codepoint (DSCP) value between "0" and "63". (for additional details see RFC2474 "Definition of the Differentiated Services Field in the IPv4 and IPv6 Headers"). Layer 3 devices that are DS compliant ("DS nodes") apply a particular per-hop forwarding behavior to data packets based on the contents of their DS fields 132. This mechanism
20 provides a method for dividing or allocating bandwidth of a network between the different flows, and is generally referred to as the per-hop-behavior (PHB). Examples of per-hop forwarding behaviors include expedited forwarding (EF) and assured forwarding (AF). Additional information on AF and EF forwarding can be found in RFC2597 and RFC2598.

25 In a typical differential services environment, DS nodes located at the border of the DS domain ("edge devices") mark or "color" each IP packet for a particular flow with a specific DSCP value based on the currently established QoS policies. Such coloring

may involve loading the DS field 132 of a packet with a particular DSCP value.

Thereafter, the interior DS compliant devices along the path apply the corresponding forwarding behavior to the packet based on the particular DSCP value.

For example, a QoS policy typically includes a filter or Boolean expression that indicates which packets are to be colored, and with what DSCP values. Conventionally, a network administrator selects one or more QoS policies based on a predetermined priority factor. For example, a network administrator may select a QoS policy that colors all Voice Over IP (VOIP) packets with a high priority DSCP value (for example “60”) but which marks all email packets with a low priority DSCP value (for example “10”). Thus, various DSCP values may be associated with various services.

DEFICIENCIES OF PAST APPROACHES

In general, the network administrator is responsible for defining the QoS provided within a network. Conventionally, to achieve a specified per-hop-behavior for a particular QoS, a policy management station instructs DS-compliant network nodes that are within its management domain to color each packet for a particular flow with a static DSCP value. Thereafter, the DS nodes within the DS domain forward the packets through the network based on packet’s color. However, a drawback with coloring the packets for a particular flow with static DSCP values is that the per-hop-behavior that is applied to each flow does not take into account the dynamic state of the network. In particular, the then-current loading or available bandwidth of the network is not taken into account. Thus, the coloring of packets based on the static DSCP values can reduce the throughput or bandwidth that is achieved within a DS domain as unutilized bandwidth cannot be shared among the different flows. As a result, network performance suffers.

For example, an online trading application that generates stock quote messages may be assigned a DSCP value “60” that provides a target bandwidth of fifty percent (50%). Additionally, the corporate financial information messages may be assigned a DSCP value of “45” that provides a target bandwidth of forty percent (40%), while the

data backup messages may be assigned a DSCP value of "15" that provides a target bandwidth of ten percent (10%). Thus, regardless of the current bandwidth that is available in the network for each particular flow, all stock quote message packets are colored with a DSCP value of "60", all corporate financial information message packets are colored with a DSCP value of "45", all data backup message packets are colored with a DSCP value of "15". Therefore, even if the network has more available bandwidth (for example because the network is not currently routing any stock quote messages), all incoming data backup message packets will be colored with the static DSCP value of "15" and thus not take advantage of the available unused bandwidth.

Moreover, the bandwidth that is allocated for each color (DSCP value) must be shared between all flows having the same color (i.e., assigned the same "Service Level"). Thus, the actual forwarding behavior that is applied to a particular flow is determined not only by the color of a particular flow (i.e., GOLD, SILVER, BRONZE, etc.), but also by the number of active flows that are currently associated with the same color (i.e., having packets with the same DSCP value).

For example, all VOIP flows may be associated with the color GOLD so that they are allocated fifty percent (50%) of the network bandwidth while all email flows may be associated with the color Bronze so that they are allocated twenty percent (20%) of the network bandwidth. However, if five VOIP flows are currently active in the network, each flow will allocated approximately ten percent (50% divided by 5 equals 10%) of the network bandwidth. In addition, if there is only one email flow that is currently active in the network, the email flow will allocated twenty percent (20% divided by 1 equals 20%) of the network bandwidth. Thus, in certain situations, a flow that is colored GOLD may actually receive less network bandwidth than a flow that is colored BRONZE.

Based on the foregoing, there is a clear need for a mechanism that can take advantage of unused network bandwidth in applying QoS within a DS domain.

There is also a need for a mechanism that takes into account the actual traffic load of the network at the time packets are colored within a DS domain.

- In addition, there is also a need for mechanism that can provide a higher degree of control and predictability of the traffic patterns that exist within a network when applying
- 5 a QoS within a DS domain.

SUMMARY OF THE INVENTION

According to one aspect of the invention, a method for performing load-based packet marking within a network. In this aspect, a first group of one or more packets of a data flow are marked with a first behavioral treatment value that directs devices within the network to treat the first group of one or more packets with a first quality of service treatment. The bandwidth that is currently being achieved for the flow within the network is determined based on data traffic within the network. Based on the achieved flow bandwidth within the network a second behavioral treatment value is then determined. Thereafter, a second group of one or more packets of the data flow is marked with a second behavioral treatment value that directs devices within the network to treat the second group of one or more packets with a second quality of service treatment. This process of dynamically marking the packets for a particular data flow may be performed multiple times.

According to one feature of this aspect, the step of marking a first group of one or more packets includes the step of storing a first differentiated services codepoint (DSCP) value in each header of the first group of one or more packets of a data flow; the step of determining a second behavioral treatment value includes the step of determining a second DSCP value; and the step of marking a second group of one or more packets includes the step of storing the second DSCP value in each header of the second group of one or more packets of a data flow.

According to another aspect, packet flow characteristics of the first group of one or more packets of a data flow are determined. The second behavioral treatment value is then based on both the available bandwidth within the network and the packet flow characteristics of the first group of one or more packets of a data flow.

The invention also encompasses a computer-readable medium, a computer data signal embodied in a carrier wave, and an apparatus configured to carry out the foregoing steps. Other features and aspects will become apparent from the following description and the appended claims.

006260" 08652960

BRIEF DESCRIPTION OF THE DRAWINGS

The present invention is illustrated by way of example, and not by way of limitation, in the figures of the accompanying drawings and in which like reference numerals refer to similar elements and in which:

5 FIG. 1A is a partial block diagram of a network message.

FIG. 1B is a partial block diagram of a network message.

FIG. 1C is a partial block diagram of a network message.

FIG. 2 is a block diagram of a computer network in which in which the present invention may be utilized.

10 FIG. 3 is a flow diagram that illustrates a method for performing load-based packet marking.

FIG. 4 is a flow diagram that illustrates additional details for dynamically coloring packets within a flow based on an estimated bandwidth that is currently available within a network system.

15 FIG. 5 is a block diagram of a computer system with which an embodiment may be carried out.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

A method and apparatus for marking network data packets with differentiated services codepoints based on network load is provided. In the following description, for the purposes of explanation, numerous specific details are set forth in order to provide a
5 thorough understanding of the present invention. It will be apparent, however, to one skilled in the art that the present invention may be practiced without these specific details. In other instances, well-known structures and devices are shown in block diagram form in order to avoid unnecessarily obscuring the present invention.

OPERATIONAL CONTEXT

10 A load-based packet marking mechanism is provided for dynamically coloring packets of a flow based on the bandwidth that is that is actually being achieved by the flow within a network ("achieved flow bandwidth"). In one embodiment, one or more packets for a particular flow within a network are colored using an initial marking value. Thereafter, the traffic flow of the network is analyzed to determine an achieved flow
15 bandwidth for the particular flow. Based on the achieved flow bandwidth, a new marking value is calculated and used for coloring subsequent packets within the particular flow. As will become apparent by the description below, embodiments may encompass any communication protocol. As such, the dynamic coloring of packets based on an estimate of the achieved flow bandwidth may be applied to a variety of different communication
20 protocols. These protocols may include, but are not limited to, TCP, UDP, and RTP, in addition to a variety of other communication protocols that may be used for communicating between devices over a network.

FIG. 2 is a block diagram of a computer network 200 illustrating certain elements of an embodiment. Generally, computer network 200 includes one or more network
25 devices 202, 204, a plurality of workstations 210, 212, a policy management station 208 and a network 206.

006260-08652960

Network devices 202, 204 represent network edge devices such as routers, switches, or other similar or equivalent devices that are configured for coloring packets within network 206. In one embodiment, network devices 202, 204 are configured to execute the Cisco Internetworking Operating System (IOS) and capable of marking
5 packets with DSCP values, i.e., they are compatible with Differentiated Services. Such marking may be carried out using a marker or other software element or application that runs under control of IOS, e.g., an agent or process.

Workstations 210, 212 may be personal computers, workstations, or other network end stations at which work is done, such as printers, scanners, facsimile
10 machines, etc. In certain embodiments, workstations 210, 212 may themselves be network devices, such as bridges, gateways, routers or switches that allow computer network 200 to connect to another network system. For example, workstation 210 may be an edge device that is configured for coloring packet of a different DS domain.

Network 206 is a network system comprising any number of network devices.
15 Network 206 may form part of a LAN or WAN. In one embodiment, network 206 is a packet-switched IP network configured as a DS domain whereby treatment of packets that flow through network 206 is controlled and managed by Policy Management Station 208 and network devices 202, 204.

Policy Management Station 208 is a computer, or a group of hardware or software
20 components or processes that cooperate or execute in one or more computer systems. As illustrated, Policy Management Station 208 is coupled to the one or more network devices 202, 204. Policy Management Station 208 configures the network devices 202, 204 to control the coloring of packets within network 206. For example, Policy Management Station 208 may direct network device 212 to color all Voice Over IP (VOIP) packets for
25 a particular flow with the color gold (high priority) and all File Transfer Protocol (FTP) packets of a particular flow with the color Bronze (low priority). Each color is associated with one or more QoS treatment parameters, e.g., a pre-defined DSCP value and possibly

other values or characteristics. Although not depicted in FIG. 2, Policy Management Station 208 may actually be coupled to and thus communicate with the one or more network devices 202, 204 via network 206.

In one embodiment, Policy Management Station 208 provides a mechanism whereby a network administrator may define an initial QoS treatment that is applied to the different services and flows within network 206. For example, an administrator may define an initial QoS treatment for a particular flow by setting an initial DiffServ Codepoint value in the packets, by setting IP Precedence values in the packets, or by signaling using RSVP. An example of a commercial product suitable for use as Policy Management Station 208 is CiscoAssure QoS Policy Manager 1.0, commercially available from Cisco Systems, Inc.

Estimator 214 is one or more hardware or software elements that are configured for estimating the real-time runtime bandwidth that is achieved by the different flows within network 206. In one embodiment, estimator 214 communicates with devices 202, 204 to estimate the throughput, network latency, and other packet flow characteristics that are currently associated with flows within network 206. Based on these estimates, the estimator 214 determines the bandwidth that is achieved for the different flows within network 206. Policy Management Station 208 then communicates with markers or other elements of devices 202, 204 to dynamically color the packets for each flow based on the currently achieved flow bandwidth within network 206 and the bandwidth needs for each particular flow. Examples of how the achieved flow bandwidth may be determined are described in detail below.

Although the example embodiment of FIG. 2 shows two (2) workstations 210, 212 and two (2) devices 202, 204, in other practical embodiments there may be any number of such elements.

LOAD-BASED PACKET MARKING

FIG. 3 is a flow diagram 300 that illustrates a method for performing load-based packet marking in accordance with certain embodiments of the invention. For explanation purposes, the blocks of FIG. 3 are described in reference to the components of FIG. 2. However, embodiments of the methods disclosed herein are not limited to the example embodiment that is shown in FIG. 2.

At block 302, a first QoS value is defined for coloring packets within a data flow. Block 302 also may involve defining QoS values for a plurality of flows. For example, by interfacing with Policy Management Station 208 an administrator may define a QoS policy whereby VoIP packets for flow A are colored GOLD for high forwarding priority (for example allocating 60% of the network bandwidth) and FTP packets for flow B are colored BRONZE for low forwarding priority (for example allocating 20% of the network bandwidth).

At block 304, a first group of one or more packets of the data flow with the first QoS value are colored. Block 304 may involve communicating the initial QoS values to each of the edge devices so that the devices can use the values to color the packets that are associated with each of the flows. For example, in response to an administrator defining a QoS policy that colors VOIP packets for flow A to GOLD and FTP packets for flow B to BRONZE, the Policy Management Station 208 communicates the initial QoS values to edge devices 202, 204. Thereafter, edge devices 202, 204 begin applying the QoS policies by coloring the packets of each flow based on the initial QoS values.

At block 306, traffic bandwidth within the network is estimated. As described detail below, several techniques may be used to determine the achieved flow bandwidth that exists within a network. In one embodiment, estimator 214 communicates with devices 202, 204 to collect bandwidth information corresponding to the traffic patterns that currently exist in network 206.

At block 308, one or more updated QoS values are determined for coloring packets within the data flows, based on the estimated network traffic bandwidth. For example, if estimator 214 determines that the VoIP packets for flow A are consuming only 50 percent of their allocated bandwidth (i.e., 30% of the network bandwidth) and that the FTP packets of flow B are consuming 100 percent of their allocated bandwidth (i.e., 20% of the network bandwidth), estimator 214 may determine that the FTP packets of flow B should now be colored SILVER so as to have a medium forwarding priority (for example allocating 40% of the network bandwidth). In one embodiment, the estimator 214 selects the updated QoS values based on the minimal amount of service level that is required for efficiently servicing the packets for each flow. For example, if the estimator 214 determines that a third of the FTP packets for flow B are not being efficiently forwarded through the network based on the initial QoS value, the estimator 214 can generate an updated QoS value that provides for an additional 33% bandwidth for the FTP packets for flow B.

At block 310, a second group of one or more packets of the data flow are colored with the one or more updated QoS values. The updated QoS values are communicated to each of the edge devices. Thereafter, the edge devices proceed to color the packets that are associated with each of the flows based on the updated QoS values. For example, in response to communicating the updated QoS values to edge devices 202, 204, edge devices 202, 204 begin coloring the FTP packets of flow B with the color SILVER.

As illustrated in FIG. 3 by path 312, the process of dynamically updating the coloring of packets within a particular flow based on the achieved flow bandwidth can be repeated performed to continually tune the network to enhancing efficiency of a network system.

ESTIMATING CURRENT ACHIEVED FLOW BANDWIDTH

FIG. 4 is a flow diagram that illustrates an example of a method for dynamically coloring packets within a flow based on an estimated bandwidth that is currently being

achieved by a flow within a network system. As depicted by step 402, one or more QoS policies are defined that specify target bandwidth values and a range of possible services for each flow within the network. In one embodiment, a target bandwidth identifies a specific bandwidth that is desirous or required by a particular flow within the network.

5 At step 404, an initial DSCP value for the packets of each flow is determined based on the defined QoS policies. The initial DSCP values are then communicated to the edge devices and initially used by the edge devices to color the packet of each flow.

 At block 416, information is gathered about the currently achieved flow bandwidth within the network. Several methods may be used to gather the achieved flow bandwidth information and embodiments of the invention encompass numerous methods for gathering the achieved flow bandwidth information. For example, as depicted in steps 10 406, 408, 410, certain flows may contain TCP packets (TCP flow), UDP or RTP packets (RTP flow) or a variety of other special protocol packets (Special flow). In one embodiment, MIB variables are used to estimate the achieved flow bandwidth for a particular flow.

 Alternatively, or in addition, other mechanisms may be used to gather current achieved flow bandwidth information for a particular flow. For example, the TCP/IP window size may be checked to determine the achieved flow bandwidth based on the initial DSCP value that is associated with the TCP flow. Based on the TCP/IP window 15 size, the estimator 214 may update the DSCP value of the TCP flow accordingly based on the achieved flow bandwidth.

 As another example, for UDP packets, the control flow protocol of the Real-Time Protocol (RTCP) may be used to determine the bandwidth that is currently being achieved by a particular flow. For example, RTP receivers may be used to provide reception 20 quality feedback using RTCP report packets (sender report (SR) and receiver report (RR)). Each reception report block provides statistics about the data received from the particular source. Cumulative counts are used in both the sender information and receiver

report blocks so that differences may be calculated between any two reports to make measurements over both short and long time periods, and to provide resilience against the loss of a report. The difference between the last two reports that were received can be used to estimate the recent quality of the distribution. In one embodiment, the Network Time Protocol (NTP) timestamp is included so that rates may be calculated from these differences over the interval between two reports. The time stamp shows the wall-clock time (absolute time) which is represented using the time stamp format of the Network Time Protocol (see RFC1305). Because the timestamp is independent of the clock rate for the data encoding, it is possible to implement encoding and profile independent quality monitors and to determine the achieved bandwidth for a particular flow. Additional information regarding the RTCP may be found in RFC1889.

At step 412, the achieved flow bandwidth for one or more flows is determined based on the gathered bandwidth information.

At step 414, updated DSCP values are determined based on the bandwidth estimates. The updated DSCP values are then applied to the corresponding packets of each flow. This process may be repeated one or more times to periodically color packets of a flow based on bandwidth that is currently being achieved for a particular flow within the network.

HARDWARE EXAMPLE

FIG. 5 is a block diagram that illustrates a computer system 500 upon which an embodiment of the invention may be implemented. Computer system 500 includes a bus 502 or other communication mechanism for communicating information, and a processor 504 coupled with bus 502 for processing information. Computer system 500 also includes a main memory 506, such as a random access memory (RAM) or other dynamic storage device, coupled to bus 502 for storing information and instructions to be executed by processor 504. Main memory 506 also may be used for storing temporary variables or other intermediate information during execution of instructions to be executed by

processor 504. Computer system 500 further includes a read only memory (ROM) 508 or other static storage device coupled to bus 502 for storing static information and instructions for processor 504. A storage device 510, such as a magnetic disk or optical disk, is provided and coupled to bus 502 for storing information and instructions.

5 Computer system 500 may be coupled via bus 502 to a display 512, such as a cathode ray tube (CRT), for displaying information to a computer user. An input device 514, including alphanumeric and other keys, is coupled to bus 502 for communicating information and command selections to processor 504. Another type of user input device is cursor control 516, such as a mouse, a trackball, or cursor direction keys for
10 communicating direction information and command selections to processor 504 and for controlling cursor movement on display 512. This input device typically has two degrees of freedom in two axes, a first axis (e.g., x) and a second axis (e.g., y), that allows the device to specify positions in a plane.

The invention is related to the use of computer system 500 for performing load-
15 based packet marking. According to one embodiment of the invention, load-based packet marking is provided by computer system 500 in response to processor 504 executing one or more sequences of one or more instructions contained in main memory 506. Such instructions may be read into main memory 506 from another computer-readable medium, such as storage device 510. Execution of the sequences of instructions
20 contained in main memory 506 causes processor 504 to perform the process steps described herein. One or more processors in a multi-processing arrangement may also be employed to execute the sequences of instructions contained in main memory 506. In alternative embodiments, hard-wired circuitry may be used in place of or in combination with software instructions to implement the invention. Thus, embodiments of the
25 invention are not limited to any specific combination of hardware circuitry and software.

The term "computer-readable medium" as used herein refers to any medium that participates in providing instructions to processor 504 for execution. Such a medium may

take many forms, including but not limited to, non-volatile media, volatile media, and transmission media. Non-volatile media includes, for example, optical or magnetic disks, such as storage device 510. Volatile media includes dynamic memory, such as main memory 506. Transmission media includes coaxial cables, copper wire and fiber optics, including the wires that comprise bus 502. Transmission media can also take the form of acoustic or light waves, such as those generated during radio wave and infrared data communications.

Common forms of computer-readable media include, for example, a floppy disk, a flexible disk, hard disk, magnetic tape, or any other magnetic medium, a CD-ROM, any other optical medium, punch cards, paper tape, any other physical medium with patterns of holes, a RAM, a PROM, and EPROM, a FLASH-EPROM, any other memory chip or cartridge, a carrier wave as described hereinafter, or any other medium from which a computer can read.

Various forms of computer readable media may be involved in carrying one or more sequences of one or more instructions to processor 504 for execution. For example, the instructions may initially be carried on a magnetic disk of a remote computer. The remote computer can load the instructions into its dynamic memory and send the instructions over a telephone line using a modem. A modem local to computer system 500 can receive the data on the telephone line and use an infrared transmitter to convert the data to an infrared signal. An infrared detector coupled to bus 502 can receive the data carried in the infrared signal and place the data on bus 502. Bus 502 carries the data to main memory 506, from which processor 504 retrieves and executes the instructions. The instructions received by main memory 506 may optionally be stored on storage device 510 either before or after execution by processor 504.

Computer system 500 also includes a communication interface 518 coupled to bus 502. Communication interface 518 provides a two-way data communication coupling to a network link 520 that is connected to a local network 522. For example,

communication interface 518 may be an integrated services digital network (ISDN) card or a modem to provide a data communication connection to a corresponding type of telephone line. As another example, communication interface 518 may be a local area network (LAN) card to provide a data communication connection to a compatible LAN.

5 Wireless links may also be implemented. In any such implementation, communication interface 518 sends and receives electrical, electromagnetic or optical signals that carry digital data streams representing various types of information.

Network link 520 typically provides data communication through one or more networks to other data devices. For example, network link 520 may provide a connection

10 through local network 522 to a host computer 524 or to data equipment operated by an Internet Service Provider (ISP) 526. ISP 526 in turn provides data communication services through the worldwide packet data communication network now commonly referred to as the "Internet" 528. Local network 522 and Internet 528 both use electrical, electromagnetic or optical signals that carry digital data streams. The signals through the

15 various networks and the signals on network link 520 and through communication interface 518, which carry the digital data to and from computer system 500, are exemplary forms of carrier waves transporting the information.

Computer system 500 can send messages and receive data, including program code, through the network(s), network link 520 and communication interface 518. In the

20 Internet example, a server 530 might transmit a requested code for an application program through Internet 528, ISP 526, local network 522 and communication interface 518. In accordance with the invention, one or more such downloaded applications provide for performing load-based packet marking as described herein.

The received code may be executed by processor 504 as it is received, and/or

25 stored in storage device 510, or other non-volatile storage for later execution. In this manner, computer system 500 may obtain application code in the form of a carrier wave.

ALTERNATIVES, EXTENSIONS

006260-08652960

The described load-based packet marking mechanism provides for the dynamic coloring of packets for a particular flow based on the bandwidth that is currently being achieved by the particular flow within a network. By determining the current bandwidth that is actually available for a particular flow within the network, packets for the particular flow may be dynamically colored based on the current network load. Thus, the network can perform more efficiently as available bandwidth can be dynamically distributed to the different flows within the network. In one aspect, packet flow characteristics are used to identify data flows that would benefit from receiving additional bandwidth within the network and to identify data flows whose current bandwidth allocation may be reduced, at least temporarily.

In describing certain embodiments of the invention, several drawing figures have been used for explanation purposes. However, the invention is not limited to any particular context as shown in drawing figures, and the spirit and scope of the invention include other contexts and applications in which the mechanisms and processes described herein is available to other mechanisms, methods, programs, and processes. Thus, the specification and drawings are, accordingly, to be regarded in an illustrative rather than a restrictive sense.

For example, although FIG. 2 illustrates a single policy management station 208, in certain embodiments, multiple policy management station 208 can be used for dynamically coloring packets of a particular flow based on the bandwidth that is currently being achieved by the flow within the network system.

In addition, in this disclosure, including in the claims, certain process steps are set forth in a particular order, and alphabetic and alphanumeric labels are used to identify certain steps. Unless specifically stated in the disclosure, embodiments of the invention are not limited to any particular order of carrying out such steps. In particular, the labels

are used merely for convenient identification of steps, and are not intended to imply, specify or require a particular order of carrying out such steps.

005250" 08632960

CLAIMS

What is claimed is:

- 1 1. A method for marking one or more packets of data in a packet-switched network
2 based on achieved flow bandwidth information within the network, comprising the
3 computer-implemented steps of:
4 marking a first group of one or more packets of a data flow with a first behavioral
5 treatment value, wherein the first behavioral treatment value directs
6 devices within the network to treat the first group of one or more packets
7 with a first quality of service treatment;
8 determining an achieved flow bandwidth for the data flow based on data traffic
9 within the network;
10 determining a second behavioral treatment value based on the achieved flow
11 bandwidth within the network; and
12 marking a second group of one or more packets of said data flow with said second
13 behavioral treatment value, wherein the second behavioral treatment value
14 directs devices within the network to treat the second group of one or more
15 packets with a second quality of service treatment.
- 1 2. The method as recited in Claim 1, wherein:
2 the step of marking a first group of one or more packets includes the step of
3 storing a first differentiated services codepoint (DSCP) value in each
4 header of the first group of one or more packets of a data flow;
5 the step of determining a second behavioral treatment value includes the step of
6 determining a second DSCP value; and
7 the step of marking a second group of one or more packets includes the step of
8 storing the second DSCP value in each header of the second group of one
9 or more packets of a data flow.

- 1 3. The method as recited in Claim 1, further comprising the steps of:
2 determining packet flow characteristics of the first group of one or more packets
3 of a data flow; and
4 determining the second behavioral treatment value based on the available
5 bandwidth within the network and the packet flow characteristics of the
6 first group of one or more packets of a data flow.
- 1 4. The method as recited in Claim 1, further comprising the steps of:
2 establishing a Quality of Service (QoS) policy for applying a per-hop-behavior
3 treatment for forwarding packets within a flow in said network; and
4 generating the first behavioral treatment value based on the established QoS
5 policy.
- 1 5. A computer-readable medium carrying one or more sequences of instructions for
2 marking one or more packets of data in a packet-switched network based on
3 achieved flow bandwidth information within the network, wherein execution of
4 the one or more sequences of instructions by one or more processors causes the
5 one or more processors to perform the steps of:
6 marking a first group of one or more packets of a data flow with a first behavioral
7 treatment value, wherein the first behavioral treatment value directs
8 devices within the network to treat the first group of one or more packets
9 with a first quality of service treatment;
10 determining an achieved flow bandwidth for the data flow based on data traffic
11 within the network;
12 determining a second behavioral treatment value based on the achieved flow
13 bandwidth within the network; and
14 marking a second group of one or more packets of said data flow with said second
15 behavioral treatment value, wherein the second behavioral treatment value

16 directs devices within the network to treat the second group of one or more
17 packets with a second quality of service treatment.

1 6. The computer-readable medium as recited in Claim 5, wherein:
2 the step of marking a first group of one or more packets includes the step of
3 storing a first differentiated services codepoint (DSCP) value in each
4 header of the first group of one or more packets of a data flow;
5 the step of determining a second behavioral treatment value includes the step of
6 determining a second DSCP value; and
7 the step of marking a second group of one or more packets includes the step of
8 storing the second DSCP value in each header of the second group of one
9 or more packets of a data flow.

1 7. The computer-readable medium as recited in Claim 5, further comprising
2 instructions for performing the steps of:
3 determining packet flow characteristics of the first group of one or more packets
4 of a data flow; and
5 determining the second behavioral treatment value based on the available
6 bandwidth within the network and the packet flow characteristics of the
7 first group of one or more packets of a data flow.

1 8. The computer-readable medium as recited in Claim 5, further comprising
2 instructions for performing the steps of:
3 establishing a Quality of Service (QoS) policy for applying a per-hop-behavior
4 treatment for forwarding packets within a flow in said network; and
5 generating the first behavioral treatment value based on the established QoS
6 policy.

1 9. A computer apparatus comprising:
2 a processor; and
3 a memory coupled to the processor, the memory containing one or more sequences
4 of instructions for marking one or more packets of data in a packet-
5 switched network based on achieved flow bandwidth information within
6 the network, wherein execution of the one or more sequences of
7 instructions by the processor causes the processor to perform the steps of:
8 marking a first group of one or more packets of a data flow with a first
9 behavioral treatment value, wherein the first behavioral treatment
10 value directs devices within the network to treat the first group of one
11 or more packets with a first quality of service treatment;
12 determining an achieved flow bandwidth for the data flow based on data
13 traffic within the network;
14 determining a second behavioral treatment value based on the achieved
15 flow bandwidth within the network; and
16 marking a second group of one or more packets of said data flow with said
17 second behavioral treatment value, wherein the second behavioral
18 treatment value directs devices within the network to treat the second
19 group of one or more packets with a second quality of service
20 treatment.

1 10. The computer apparatus as recited in Claim 9, wherein:
2 the step of marking a first group of one or more packets includes the step of
3 storing a first differentiated services codepoint (DSCP) value in each
4 header of the first group of one or more packets of a data flow;
5 the step of determining a second behavioral treatment value includes the step of
6 determining a second DSCP value; and
7 the step of marking a second group of one or more packets includes the step of
8 storing the second DSCP value in each header of the second group of one
9 or more packets of a data flow.

- 1 11. The computer apparatus as recited in Claim 9, further comprising instructions for
2 performing the steps of:
3 determining packet flow characteristics of the first group of one or more packets
4 of a data flow; and
5 determining the second behavioral treatment value based on the available
6 bandwidth within the network and the packet flow characteristics of the
7 first group of one or more packets of a data flow.
- 1 12. The computer apparatus as recited in Claim 9, further comprising instructions for
2 performing the steps of:
3 establishing a Quality of Service (QoS) policy for applying a per-hop-behavior
4 treatment for forwarding packets within a flow in said network; and
5 generating the first behavioral treatment value based on the established QoS
6 policy.
- 1 13. A network device configured for marking one or more packets of data in a packet-
2 switched network based on achieved flow bandwidth information within the
3 network, comprising:
4 means for marking a first group of one or more packets of a data flow with a first
5 behavioral treatment value, wherein the first behavioral treatment value
6 directs devices within the network to treat the first group of one or more
7 packets with a first quality of service treatment;
8 means for determining an achieved flow bandwidth for the data flow based on data
9 traffic within the network;
10 means for determining a second behavioral treatment value based on the achieved
11 flow bandwidth within the network; and
12 means for marking a second group of one or more packets of said data flow with
13 said second behavioral treatment value, wherein the second behavioral
14 treatment value directs devices within the network to treat the second
15 group of one or more packets with a second quality of service treatment.

- 1 14. A method for marking one or more packets of data in a packet-switched network
2 based on achieved flow bandwidth information within the network, comprising the
3 computer-implemented steps of:
4 causing one or more network devices to mark a first group of one or more packets
5 of a data flow with a first behavioral treatment value, wherein the first
6 behavioral treatment value directs devices within the network to treat the
7 first group of one or more packets with a first quality of service treatment;
8 determining an achieved flow bandwidth for the data flow based on data traffic
9 within the network;
10 determining a second behavioral treatment value based on the achieved flow
11 bandwidth within the network; and
12 causing one or more network devices to mark a second group of one or more
13 packets of said data flow with said second behavioral treatment value,
14 wherein the second behavioral treatment value directs devices within the
15 network to treat the second group of one or more packets with a second
16 quality of service treatment.
17

ABSTRACT OF THE DISCLOSURE

A method and apparatus for performing load-based packet marking within a network is described. In one aspect, a first group of one or more packets of a data flow are marked with a first behavioral treatment value that directs devices within the network to treat the first group of one or more packets with a first quality of service treatment. The bandwidth that is currently being achieved for the flow within the network is determined based on data traffic within the network. Based on the achieved flow bandwidth within the network a second behavioral treatment value is then determined. Thereafter, a second group of one or more packets of the data flow is marked with a second behavioral treatment value that directs devices within the network to treat the second group of one or more packets with a second quality of service treatment. The process of dynamically marking the packets for a particular data flow may be performed multiple times.

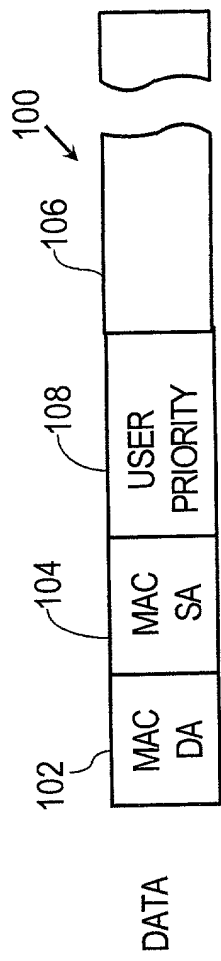


FIG. 1A (Prior Art)

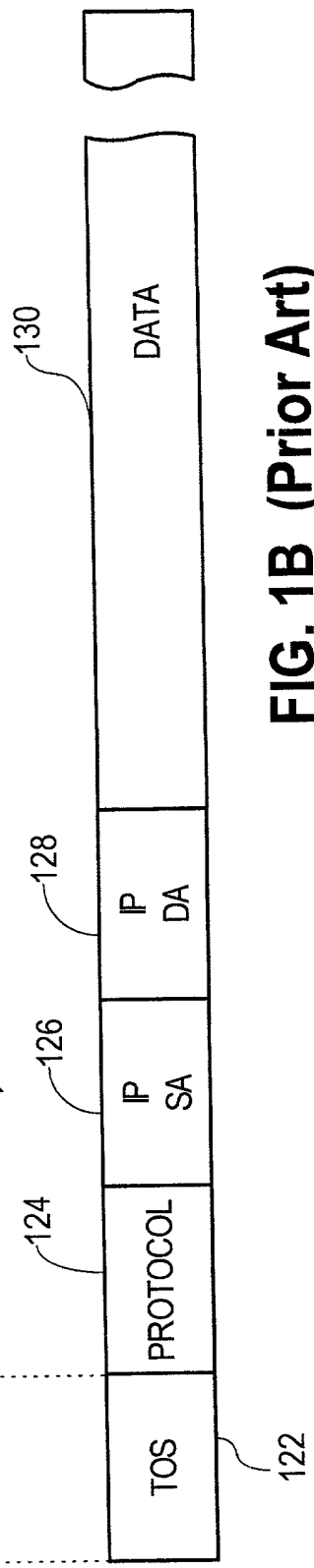


FIG. 1B (Prior Art)

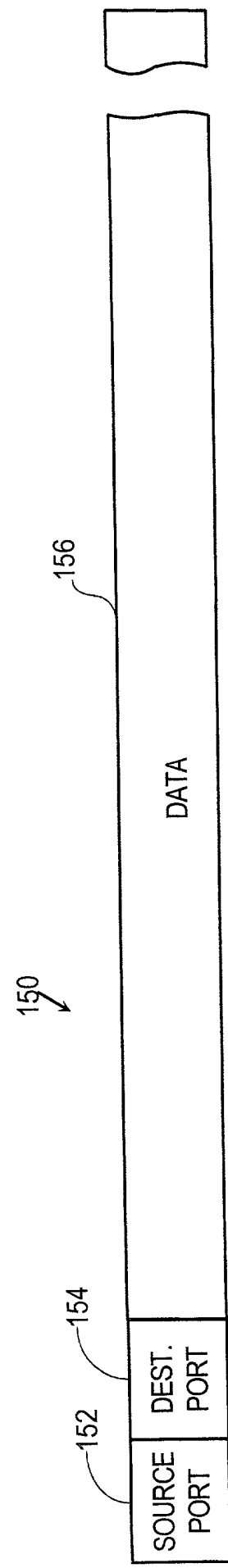


FIG. 1C (Prior Art)

200

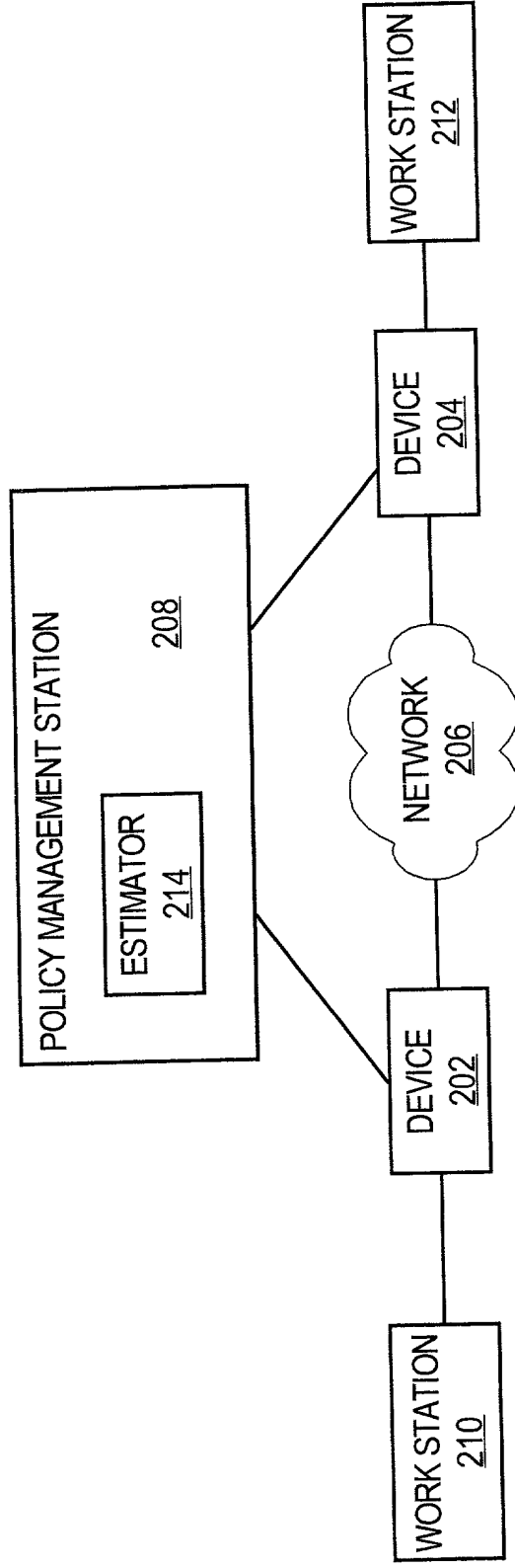


FIG. 2

006260" 08652960

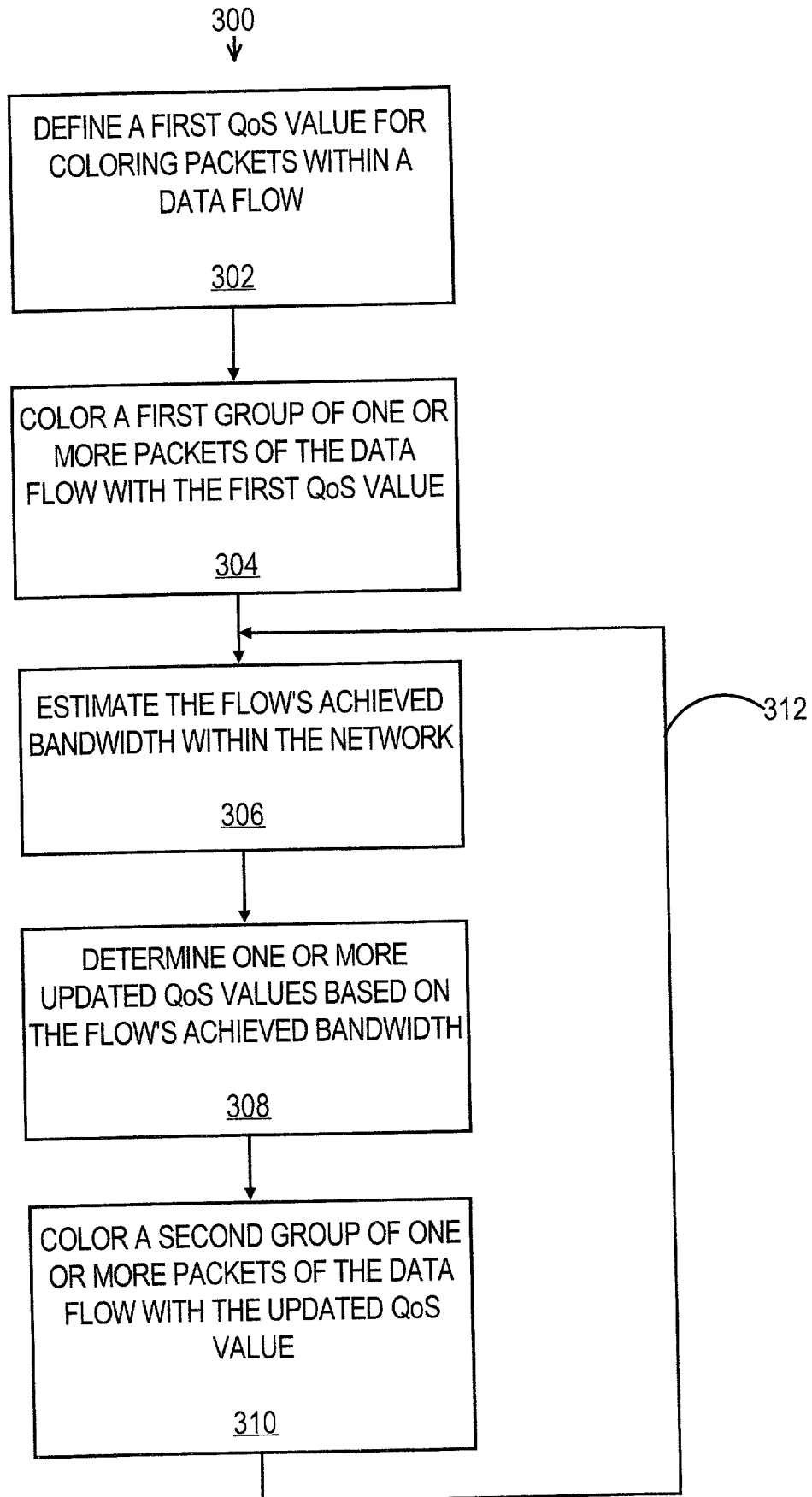


FIG. 3

FIG. 4

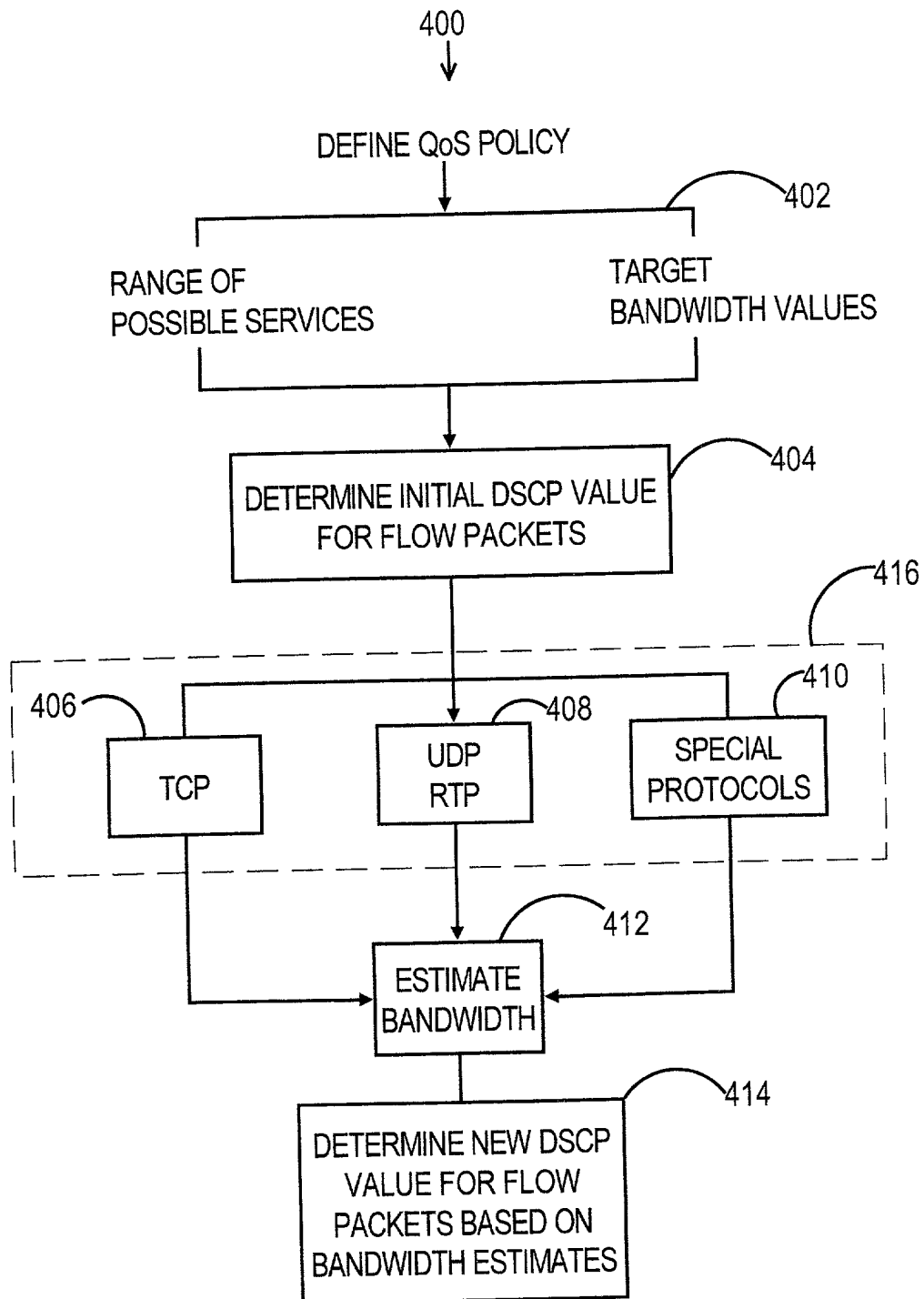
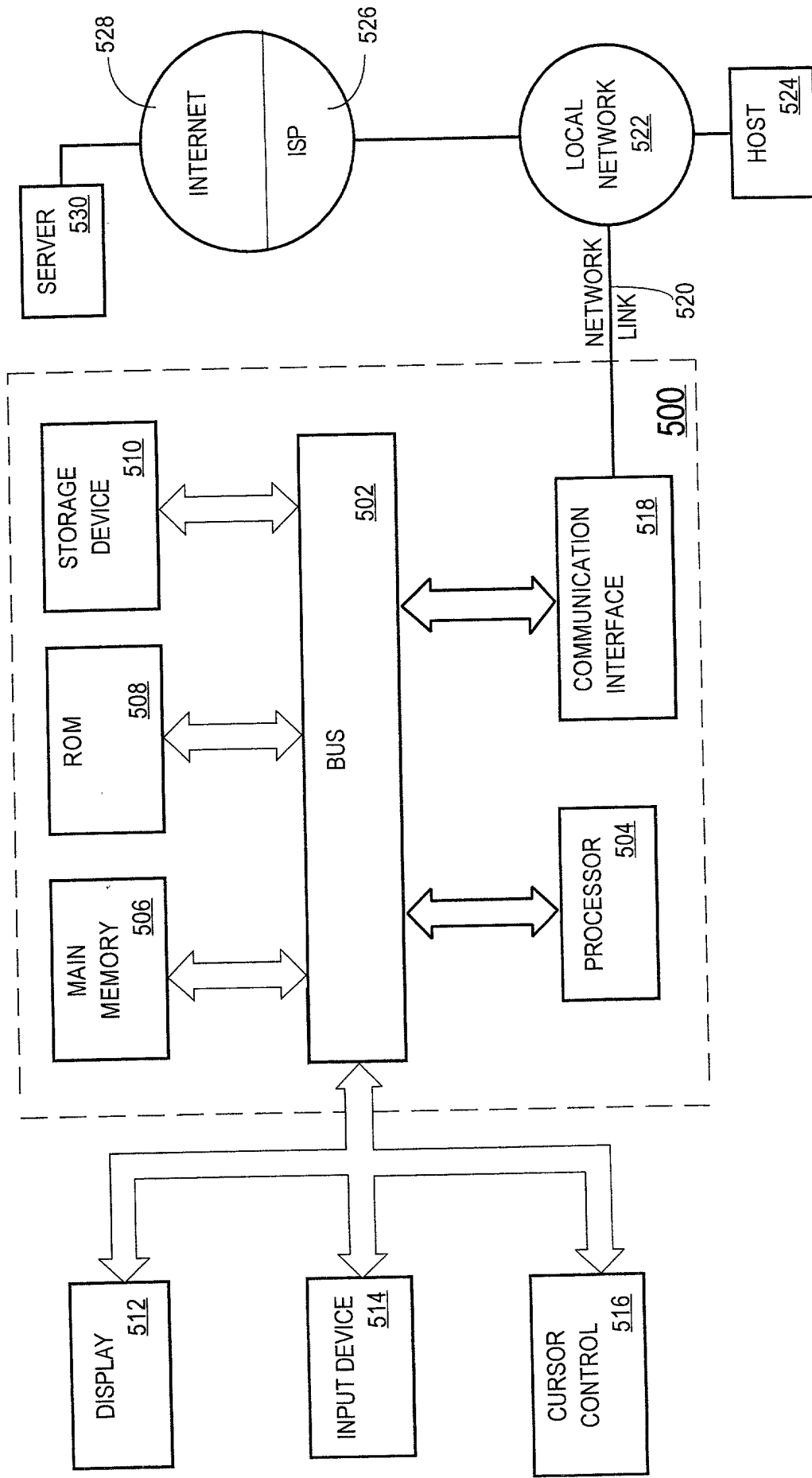


FIGURE 5



DECLARATION AND POWER OF ATTORNEY

As a below named inventor, I hereby declare that:

My residence, post office and citizenship are as stated below next to my name,

I believe I am an original, first and joint inventor of the subject matter claimed and for which a patent is sought on the invention MARKING NETWORK DATA PACKETS WITH DIFFERENTIATED SERVICES CODEPOINTS BASED ON NETWORK LOAD, the specification of which

☐ is attached hereto.
☒ was filed on _____ as Application Serial No. _____ and was amended on (if applicable).

I hereby state that I have reviewed and understand the contents of the above identified specification, including the claims, as amended by any amendment referred to above.

I acknowledge the duty to disclose information which is known to me to be material to patentability in accordance with Title 37, Code of Federal Regulations, Section 1.56.

I hereby claim foreign priority benefits under Title 35, United States Code, Section 119 of any foreign application(s) for patent or inventor's certificate listed below and have also identified below any foreign application for patent or inventor's certificate having a filing date before that of the application on which priority is claimed:

Prior Foreign Applications(s):

Number	Country	Day/Month/Year filed	Priority Claimed
			<input type="checkbox"/>
			<input type="checkbox"/>

I hereby claim the benefit under 35 USC §119(e) of any United States provisional application(s) listed below.

Prior Provisional Application(s):

Application Number	Filing Date

I hereby claim the benefit under Title 35, United States Code, Section 120 of any United States application(s) listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States application in the manner provided by the first paragraph of Title 35, United States Code, Section 112, I acknowledge the duty to disclose material information as defined in Title 37, Code of Federal Regulations, Section 1.56 which occurred between the filing date of the prior application and the national or PCT international filing date of this application:

Prior U.S. Application(s):

Serial No.

Filing Date

Status: Patented, Pending, Abandoned

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

I hereby appoint the following attorney(s) and/or agent(s): Brian D. Hickman, Reg. No. 35,894; Christopher J. Palermo, Reg. No. 42,056; Bobby K. Truong, Reg. No. 37,499; Edward A. Becker, Reg. No. 37,777; Marcel K. Bingham, Reg. No. 42,327; Carl L. Brandt, Reg. No. 44,555; and Carina M. Tan, Reg. No. 45,769, all of

HICKMAN PALERMO TRUONG & BECKER, LLP
1600 Willow Street
San Jose, California 95125-5106

with full power of substitution and revocation, to prosecute this application and to transact all business in the Patent and Trademark Office connected therewith, and all future correspondence should be addressed to them.

Full name of sole or first inventor: ARTHUR ZAVALKOVSKY

Inventor's signature:

Date:

Residence:

Citizenship:

Post Office Address:

Full name of second inventor: ITZHAK PARNAFES

Inventor's signature:

Date:

Residence: 20050 Rodrigues Avenue, Apt. B, Cupertino, California 95014

Citizenship: Israel

Post Office Address:

Docket No.: 50325-0106

Full name of third inventor: SHAI MOHABAN

Inventor's signature:

Date:

Residence: 678 Durshire Way, Sunnyvale, California 94087

Citizenship: Israel

Post Office Address:

005250-0862350